

**ISO/IEC JTC1/SC2/WG2
Coded Character Set
Secretariat: Japan (JISC)**

Doc. Type: Draft disposition of comments

Title: Draft disposition of comments on SC2 N 3914 (PDAM text for Amendment 4 to ISO/IEC 10646:2003)

Source: Michel Suignard (project editor)

Project: JTC1 02.10646.00.04

Status: For review by WG2

Date: 2007-04-22

Distribution: WG2

Reference: SC2 N3914, 3920

Medium: Paper, PDF file

Comments were received from China, Ireland, Japan, United Kingdom, and USA. The following document is the draft disposition of those comments. The disposition is organized per country. The issues concerning the CJK Unified Ideograph Extension C are consolidated in a table at the end of the document.

Note – The full content of the ballot comments have been included in this document to facilitate the reading. The dispositions are inserted in between these comments and are marked in **Underlined Bold Serif text**, with *explanatory text in italicized serif*.

China:Negative

China votes NO to ISO/IEC JTC1/SC2N3914 PDAM4 of ISO/IEC 10646: 2003. Acceptance of our comments and appropriate changes in the text will change our vote to approval.

Technical comments

T1.a. Lanna (Name of the script)

Lanna is a regional name which is named by foreign scholars for a kind of script in Tai land. Actually, the script is used internationally. It is now used or was used in not only Tai Land, but also China, Myanmar and Laos. The script is given different names in different regions although there is no remarkable differences among them. The script is even called in various names rather than Lanna in Tai Land. Thus, another name which can be accepted by all parties is needed.

WG2 decision

The Chinese Member Body already made a similar comment in document WG2 N3161, proposing then “Old Tai Lue”. However the ad-hoc during meeting WG2#50 (Tokyo, September 2006) produced document WG2 N3169 which addressed it as follows:

*“Chinese comment I. The ad-hoc group believes that the name “Lanna” is the best name for this script in the English language. The Lanna script is used to write four languages: Khün, Lao Tham, Northern Thai, and Tai Lue. The name proposed, “Old Tai Lue” is accurate in distinction to the “New Tai Lue” script already encoded, but would not be appropriate as a generic name for this script given its use in Myanmar, Thailand and Laos.”
It does not seem necessary to revisit that issue, unless a new name is effectively proposed as an alternative.*

T1.b. Characters:

The script was improved by users respectively after it was introduced to various regions in south-east Asia. In order to meet the demands of users of the script in various regions, more encoded characters are needed. For use in China, there are about 10 more consonants should be encoded.

Need further input

Cannot be accommodated unless an actual proposal is submitted. Ireland, U.K., and the U.S. are also asking for two additional Lanna characters.

T2. Mahjong characters

The following 5 glyphs should be changed to that like traditional Mahjong tiles style:



1F010 MAHJONG TILE ONE OF BAMBOOS (tiao)
1F022 MAHJONG TILE PLUM (mei)
1F023 MAHJONG TILE ORCHID (lan)
1F024 MAHJONG TILE BAMBOO (zhu)
1F025 MAHJONG TILE CHRYSANTHEMUM (ju)



WG2 decision

It is difficult to accommodate that comment without having representative glyphs submitted with the comment. The styles of Mahjong tiles vary significantly depending on sources. Note that Ireland proposes in its comment E.2 new glyphs for 1F010, 1F022, 1F023, and 1F025 (but not for 1F024). Maybe these changes would be acceptable to China?

T3 CJK Unified Ideographs Extension C: Glyph errors

The following 16 glyphs should be corrected, or should be confirmed by their submitters:

2A77D (source: V04-414D) should be  rather than .

2A825 (source: V04-4272) should be  rather than .

2A8AB (source: TE-4633) should be 聽 rather than 聽

2AB23 (source: V04-4839) should be 慍 rather than 慍

2ABB1 (source: V04-4946) should be 掣 rather than 斥

2ABF6 (source: V04-497B) should be 散酒 rather than 散酒

2AC09 (source: V04-4A25) should be 鏤 rather than 鏤

2ACF3 (source: JK-65293) should be 啄 rather than 啄

2AF72 (source: V04-4E31) should be 底生 rather than 底生

2B000 (source: V04-4F4B) should be 砗 rather than 砗

2B028 (source: V04-4F58) should be 砗 rather than 砗

2B04B (source: TD-4223) should be 襴 rather than 襴

2B0CE (source: V04-5035) should be 筭 rather than 筭

2B187 (source: V04-5142) should be 羴 rather than 羴

2B3BF (source: V04-544F) should be 蒔 rather than 蒔

2B70A (source: JK-66061) should be 鵠 rather than 鵠

Propose acceptance

Note that errors for 2ABB1 and 2B0CE were also captured by UK (comments T.3 and T.4). Japan also noticed some of these errors and then more. See table at end of document.

T.4 CJK Unified Ideographs Extension C: last digit of some code points missing:

2AACA	2AACB	2AACC	2AACD	2AACE	2AACF
2AADA	2AADB	2AADC	2AADD	2AADE	2AADF
2AAEA	2AAEB	2AAEC	2AAED	2AAEE	
2AAFA	2AAFB	2AAFC	2AAFD		
2ABAA	2ABAB	2ABAC	2ABAD	2ABAE	2ABAF
2ABBB	2ABBB	2ABBC	2ABBD	2ABBE	2ABBF

2ABCA	2ABCB	2ABCC	2ABCD	2ABCE	2ABCF
2ABDA	2ABDB	2ABDC	2ABDD	2ABDE	2ABDF
2ABEA	2ABEB	2ABEC	2ABED	2ABEE	
2ABFA	2ABFB	2ABFC	2ABFD		
2ACAA	2ACAB	2ACAC	2ACAD	2ACAE	2ACAF
2ACBA	2ACBB	2ACBC	2ACBD	2ACBE	2ACBF
2ACCA	2ACCB	2ACCC	2ACCD	2ACCE	2ACCF
2ACDA	2ACDB	2ACDC	2ACDD	2ACDE	2ACDF
2ACEA	2ACEB	2ACEC	2ACED	2ACEE	
2ACFA	2ACFB	2ACFC	2ACFD		
2ADAA	2ADAB	2ADAC	2ADAD	2ADAE	2ADAF
2ADBA	2ADBB	2ADBC	2ADBD	2ADBE	2ADBF
2ADCA	2ADCB	2ADCC	2ADCD	2ADCE	2ADCF
2ADDA	2ADDB	2ADDC	2ADDD	2ADDE	2ADDF
2ADEA	2ADEB	2ADEC	2ADED	2ADEE	
2ADFA	2ADFB	2ADFC	2ADFD		
2AECA	2AECB	2AECC	2AECD	2AECE	
2AEDA	2AEDB	2AEDC	2AEDD	2AEDE	
2AFAA	2AFAB	2AFAC	2AFAD		
2AFBA	2AFBB	2AFBC	2AFBD		
2AFCA	2AFCB	2AFCC	2AFCD		
2AFDA	2AFDB	2AFDC	2AFDD		

Propose acceptance

In fact, as noted by Japan in their attachment and UK in comment E.4, even more entries are defective. This is probably a formatting error during the chart printing process. This will be fixed in the next phase of this amendment. The source file (CJKU_Sr.txt) has the correct values.

Ireland: Negative

Ireland **disapproves** the draft with the technical and editorial comments given below.
Acceptance of these comments and appropriate changes to the text will change our vote to approval.

Technical comments

Page 14, Table 62 - Row 1A: Lanna.

With reference to ISO/IEC JTC1/SC2/WG2 N3207 “Revised proposal for encoding the Lanna script in the BMP of the UCS”, Ireland requests that the characters 𑄛 LANNA LETTER KHUN HIGH CHA and 𑄛 LANNA SIGN CAANG be added to the PDAM at positions U+1A29 (moving the following characters down one position to U+1A5F) and U+1AAD respectively. Ireland also requests that the character names as given in N3207 be used for all characters, replacing the names used in the PDAM. (This will correct some errors arising from the PDAM names apparently being taken from N3121 instead of from N3121R.)

Propose acceptance

However the UK in its comment T1 has additional renaming requests that need to be merged with this comment.

T2. Page 18, Table 110 - Row 2D: Cyrillic Extended-A.

With reference to ISO/IEC JTC1/SC2/WG2 N3194 “Proposal to encode additional Cyrillic characters in the BMP of the UCS”, Ireland requests that the character names and glyphs for the characters listed in this chart be changed to those shown on pages 16 and 17 of N3194. The names should reflect the UCS names for the Cyrillic base letters, not the Church Slavic names for these characters. The glyphs should be in Roman style (graẓdanka), not in Slavonic style. This change will also affect the characters 2DE0 .. 2DF5 in the “List of combining characters” on page 6 of the PDAM.

Regarding the other Cyrillic characters proposed in N3194, Ireland would favour the addition of these to the PDAM, but does not make this a condition for changing our vote to approval.

Propose partial acceptance

Similar request from the US (comment T.3) concerning renaming. The additional characters should be dealt outside of the context of the disposition of comments, given the very large size of the addition request.

Editorial comments

E1. Page 22, Table 201 - Row 01: Ancient Symbols.

Ireland requests that the winding error in the glyph for U+10194 be corrected. We would also like clarification: should the header of the table not be listed as “Row 101”?

Proposed acceptance in principle

The glyph correction is fine. Concerning Row numbers, it is always within a plane. The chart mentions P=1, which is the indicator that the characters belongs to Plane 1.

E2. Page 24, Table 236 - Row F0: Mahjong Tiles.

The glyphs for U+1F010 MAHJONG TILE ONE OF BAMBOOS, U+1F022 MAHJONG TILE PLUM, U+1F023 MAHJONG TILE ORCHID, and U+1F025 MAHJONG TILE CHRYSANTHEMUM, should be improved. We propose the following glyphs (showing the tile for U+1F024 MAHJONG TILE BAMBOO for comparison with the other flowers):



The bird represents a sparrow, or Chinese máquè, another name for májiàng ‘mahjong’. We would also like clarification: should the header of the table not be listed as “Row 1F0”?

Accepted in principle

Note similar request from China (T.2), although China also asks for the replacement of the representative glyph for 1F024 but without giving an example. Concerning the row number, see the disposition of comment E.1.

Japan, Negative

Japan disapproves the document SC2 N3875 with five technical and one editorial comment. Japan will change its vote if the comments are accepted and corresponding texts are updated appropriately.

Technical comments

JP1 (Technical) Removal of implementation levels

Japan supports the idea of removing implementation levels from ISO/IEC 10646. However, the way the standard is revised to do so under the current draft is considered inappropriate. The following arrangements are suggested:

- All references to the term "implementation level" should be removed from normative text, including historical references to implementation level 3.
- Informative references to "implementation level" should also be removed as much as possible.
- The control functions for identification of coded representation forms with implementation level 3 should be redefined as identification of coded representation forms, and the phrase "with implementation level 3" should be removed from the definitions.
- Japan wants to add an informative annex to explain implementation level in the past standards, including something for control functions to identify implementation levels 1 and 2, the unique spelling rule (for Indic scripts), and collection's "automatic removal" based on designated implementation levels.

Propose acceptance in principle

When formulating his initial proposal concerning the removal of implementation level (see document WG2 N3148 section 7), the project editor considered that path of removing all the implementation level terminology from the normative text. However it is useful to put again the considerations mentioned in that document (emphasis added):

*“The proposal is to eliminate the choice of multiple implementation levels from new versions of the standard and limit implementation level to level 3. Implementations requiring the former implementation levels 1 and 2 could still refer to the standard up to ISO/IEC 10646:2003 and the two current amendments. For identification purpose, **the sole remaining level: level 3 would still be mentioned in places such as ISO/IEC 2022, to maintain compatibility with current versions of ISO/IEC 10646.** The deprecated levels would still be mentioned in notes. The following is a comprehensive set of editor’s instructions to achieve that goal:”*

The problem with totally removing implementation level is that the non restricted ‘level’ of 10646 is referenced externally with the implementation level 3. Strictly speaking, the implementation levels are not part of the escape sequences required within the context of ISO/IEC 2022; it is just an explanation in the clauses 16.2, annexes C.5 and D.6. Therefore, the level mention can be removed from these normative parts as requested by Japan.

Concerning annex N, the situation is a bit different because the level is part of the arc sequence. Therefore, for compatibility reason, the number ‘3’ must stay in the sequence, although the rationale for inclusion may be put in a note. The whole annex is informative anyway.

The document WG2 N3230 also includes proposed changes to the level description which is very similar to what the Japanese NB is seeking and is aligned with this disposition.

Finally, having a full new informative annex does not seem necessary. It would give more apparent weight to a concept we are trying to deprecate. It is always possible to refer to previous versions/editions of the standard to get more explanation about implementation levels.

JP2 (Technical) Correction to CJK Unified Ideographs Extension C

Japan found some problems in the current code chart for CJK Unified Ideographs extension C. Details of the problems are listed in the attachment. The code chart should be updated.

Japan wants WG2 to instruct IRG to evaluate the attached comments and to provide corrected code chart.

In CJK extension C code chart, V-column glyphs listed below have some differences from IRG's internal review document. Japan requests IRG to verify and confirm the shapes.

2a719, 2a71b, 2a77d, 2a825, 2a949

2aa11, 2aa84, 2ab23, 2abf6, 2ae25

2aebe, 2aec3, 2af7c, 2b000, 2b028

2b09b, 2b0ce, 2b16c, 2b187, 2b23c

2b3bf, 2b40e, 2b642, 2b644

2abb1 V-column glyph is wrong. Japan request IRG to confirm.

2b151 U-column glyph is wrong. Japan request IRG to confirm.

The chart has some printing problems. In particular, some code position values on "Ucode" column lack the last hexadecimal digit. For example, the code position "2AAAA" is printed as "2AAA". The overall margins and/or positionings should be adjust appropriately.

Propose acceptance in principle

Similar comments were made by China. What is a bit alarming is that although there is a large intersection between the discovered issues, each Member Body came with a significant amount of unique issues. It probably means that IRG should engage in a significant review of Extension C before going to the next stage of this amendment, especially concerning the V column as most of the errors concern that column. Concerning the character 2B151, the U-column shows the same glyph in the original proposal:

http://www.cse.cuhk.edu.hk/~irg/irg/irg26/IRGN1235_C1Evidence_Unicode.pdf, which itself points to the non resolvable link <http://linguistics.berkeley.edu/~rscook/bishop/Wenlin-PUA.pdf>. At first glance, the unification between 纒 and 纒 looks suspicious, maybe they should have separate code points?

Editorial comments

JP3 (Editorial) V source reference

In 27.3, change "V4-4876" to "V04-4876" in the figure to align with the actual code chart.

Propose acceptance in principle

Obviously the example should conform to the actual chart. However in this case, the chart does not use the same format for the 4th V sources (V04) than for the first 3 V sources (V1 to V3). It seems that all V source should have used the same format (V1 to V4). Therefore the preferred solution would be to have consistently formatted V sources.

United Kingdom: Negative with comments:

The UK votes to DISAPPROVE the amendment, with the following technical and editorial comments. If our comments are satisfactorily resolved we will change our vote to APPROVAL.

Technical comments:

T.1.1 Page 15 Lanna name changes

Rename the following characters. NB The changes marked with an asterisk are not reflected in N3207.

1A22 LANNA LETTER HIGH XA => LANNA LETTER HIGH KXA
1A23 LANNA LETTER LOW GA => LANNA LETTER LOW KA
1A24 LANNA LETTER LOW KHA => LANNA LETTER LOW KXA
1A25 LANNA LETTER LOW KHAA => LANNA LETTER LOW KHA
1A26 LANNA LETTER LOW NGA => LANNA LETTER NGA *
1A28 LANNA LETTER HIGH SA => LANNA LETTER HIGH CHA
1A29 LANNA LETTER LOW CHA => LANNA LETTER LOW CA
1A2A LANNA LETTER NORTHERN THAI LOW CHA => LANNA LETTER NORTHERN THAI LOW CA
1A2D LANNA LETTER LOW SAA => LANNA LETTER LOW CHA
1A2E LANNA LETTER HIGH NYA => LANNA LETTER NYA *
1A2F LANNA LETTER LATA => LANNA LETTER RATA
1A30 LANNA LETTER HIGH LATHA => LANNA LETTER HIGH RATHA
1A31 LANNA LETTER LADA => LANNA LETTER DA
1A32 LANNA LETTER LOW LATHA => LANNA LETTER LOW RATHA
1A33 LANNA LETTER LANA => LANNA LETTER RANA
1A41 LANNA LETTER LOW NYA => LANNA LETTER LOW YA
1A42 LANNA LETTER YA => LANNA LETTER HIGH YA
1A44 LANNA LETTER RU => LANNA LETTER RUE
1A46 LANNA LETTER LU => LANNA LETTER LUE
1A48 LANNA LETTER HIGH SAA => LANNA LETTER HIGH SHA
1A4A LANNA LETTER HIGH SSAA => LANNA LETTER HIGH SA
1A4C LANNA LETTER LAA => LANNA LETTER LLA
1A56 LANNA LETTER LE => LANNA LETTER LAE
1A5A LANNA SIGN KHUN MAI KANG LAI => LANNA SIGN KHUEN MAI KANG LAI *
1A5D LANNA CONSONANT SIGN HIGH LATHA OR LOW PA => LANNA CONSONANT SIGN HIGH RATHA OR LOW PA
1A6D LANNA VOWEL SIGN ONG => LANNA VOWEL SIGN O
1A6E LANNA VOWEL SIGN OH => LANNA VOWEL SIGN OA BELOW
1A71 LANNA VOWEL SIGN EE => LANNA VOWEL SIGN AE
1A75 LANNA VOWEL SIGN O => LANNA VOWEL SIGN OA ABOVE
1A79 LANNA SIGN KHUN TONE-3 => LANNA SIGN KHUEN TONE-3 *
1A7A LANNA SIGN KHUN TONE-4 => LANNA SIGN KHUEN TONE-4 *
1A7B LANNA SIGN KHUN TONE-5 => LANNA SIGN KHUEN TONE-5 *
1A7C LANNA SIGN LAHAAM => LANNA SIGN RA HAAM
1AA0 LANNA SIGN WIANGWAAK => LANNA SIGN WIANG
1AA1 LANNA SIGN WIANG => LANNA SIGN WIANGWAAK
1AA6 LANNA SIGN REVERSED ROTATED LANA => LANNA SIGN REVERSED ROTATED RANA
1AA8 LANNA SIGN GAAN => LANNA SIGN KAAAN
1AA9 LANNA SIGN GAANGUU => LANNA SIGN KAANKUU
1AAA LANNA SIGN SATGAAN => LANNA SIGN SATKAAN
1AAB LANNA SIGN SATGAANGUU => LANNA SIGN SATKAANKUU

Propose acceptance

Note that code positions reflect pdam4, not the revised values from WG2 N3207. Most of these name changes are also requested by Ireland. However the additional name changes (marked with ‘’) needed to be reviewed by WG2 experts before acceptance.*

T 1.2. Lanna code positions moves

Move the following characters, as per N3207: 1A29..1A5E => 1A2A..1A5F

Propose acceptance

Also requested by Ireland (comment T.1), and implied by US comment T.2.

T 1.3. Lanna new characters

Add the following new characters, as per N3207:

1A29 LANNA LETTER KHUEN HIGH CHA (NB suggested name differs from N3207)
1AAD LANNA SIGN CAANG

Propose acceptance

Also requested by Ireland (comment T.1), and by US comment T.2.

T 1.4. Lanna character removal

There is some doubt over the need to encode the following two characters and whether they are equivalent to the decomposition sequences assigned to them in N3207. We therefore suggest removing them from Amd.4 pending further investigation.

1A65 LANNA VOWEL SIGN AM
1A66 LANNA VOWEL SIGN TALL AM

Propose acceptance

Also requested by US comment T.2

T 1.5. Lanna code position move

Move the following characters to fill the gap left by the removal of 1A65..1A66:

1A67..1A7D => 1A65..1A7B

WG2 discussion

Moving the following characters or keeping holes depends on result of removal discussion.

T2 page 48: CJK Unified Ideographs Extension C U+2A988

(glyphs added by project editor)

We believe that according to the CJK unification rules U+2A988 嫗 should be unified with the already encoded U+2177B 嫗. The difference between U+2177B and U+2A988 is that the righthand component of U+2177B is written as U+4E8F 亏 whereas the righthand component of U+2A988 is written as U+4E90 亏. The following examples demonstrate that U+4E8F and U+4E90 are unifiable components:

U+2A746. In Amd.4 source glyph TC-4375 傍 is written with U+4E90, whereas source glyph V04-4126 傍 is written with U+4E8F.

U+28706 鄗. In Super CJK Version 14.0

<http://www.cse.cuhk.edu.hk/~irg/irg/CJK/SuperCJK140_IRGN802.zip> page 1729 U+28706 is written with both U+4E8F and U+4E90 as its righthand component.

Furthermore, ISO/IEC 10646:2003 Annex S (page 1413) gives U+6C5A 汚 (U+4E90 component) and U+6C61 污 (U+4E8F component) as an example of two characters which would have been unified according to the unification rules given (S.1.6).

We therefore request that U+2A988 be removed from the amendment, and the characters 2A989..2B77A be renumbered accordingly.

Propose acceptance

The point made is very valid. Per clause S.1.6, the source separation rule only applies to BMP ideographs (not the case here). Maybe the right decision is to move this character in a compatibility area if full mapping to source is required but anyway it does not seem to belong to a CJK unified block.

T3 page 65: CJK Unified Ideographs Extension C U+2ABB1

(glyphs added by project editor)

The glyph shown for U+2ABB1 𠬪 (V04-4946) is completely incorrect (wrong radical and five strokes instead of 14). The source glyph is also shown incorrectly in N3134A1 page 67 (#09682), although the reference image is shown correctly here 𠬪. In IRG N898 <http://www.cse.cuhk.edu.hk/~irg/irg/irg19/N898-VietNam_C1.zip> page 11 V04-4946 is shown with the correct source glyph.

We therefore request that that the glyph for U+2ABB1 be changed to reflect the source glyph V04-4946 given in the original submission by Vietnam.

Propose acceptance

Also found by Japan (comment JP2)

T4 page 106: CJK Unified Ideographs Extension C U+2B0CE

(glyphs added by project editor)

The glyph shown for U+2B0CE 筳 (V04-5035) appears to be incorrect (nine residual strokes instead of eleven). The source glyph is also shown incorrectly in N3134A2 page 140 (#16060), although the reference image is shown correctly here 筳. In IRG N898 <http://www.cse.cuhk.edu.hk/~irg/irg/irg19/N898-VietNam_C1.zip> page 18 V04-5035 is shown with the correct source glyph (eleven residual strokes rather than nine).

We therefore request that that the glyph for U+2B0CE be changed to reflect the source glyph V04-5035 given in the original submission by Vietnam.

Propose acceptance

Also found by China (comment T3) and Japan (comment JP2)

T5 page 91: CJK Unified Ideographs Extension C U+2AEEF

(glyphs added by project editor)

The glyph for U+2AEEF 獺 is the same as the glyph for the compatibility ideograph U+2F927 獺, which is canonically equivalent to U+24814 獺. The righthand component of U+2AEEF (i.e. U+8C9F 貞) is a common glyph variant of the righthand component of U+24814 (i.e. U+54E1 員), and we believe that these two components are normally unifiable. For example, ISO/IEC 10646:2003 Annex S (page 1411) gives U+570E 圓 (U+8C9F component) and U+5713 圓 (U+54E1 component) as an example of two characters which would have been unified according to the unification rules given in S.1 but for the fact that they come under the source separation rule (S.1.6).

Nevertheless, there are some examples of CJK-B characters with the U+8C9F component that do correspond to characters with the U+54E1 component (U+202CF 偵 & U+508A 偵, U+21396 填 & U+5864, U+27D80 貶 & U+27D8A 貶, U+291B9 霽 & U+291C2 霽, U+2A0F0 鷓 & U+9DB0 鷓), but in none of these cases is there a corresponding compatibility ideograph.

We therefore request clarification as to whether it is necessary and appropriate to encode U+2AEEF in addition to U+2F927.

WG2 decision

Keeping the character creates an issue for the compatibility mapping for 2F927 as strictly speaking it would need to change from 24814 to 2AEEF. However the best solution may be to remove this newly proposed character and add the source reference (G_HC100898) to 2F927.

Editorial Comments

E1. Page 4 : Page 1351, annex A.1

<quote>

In the list of collections numbers and names, after

307 UNICODE 5.0 see A6.5 *

insert the new entry:

308 UNICODE 5.1 see A6.6 *

</quote>

308 Unicode 5.1 is already defined in Amd.3. This should be:

<quote>

308 UNICODE 5.1 see A6.6 *

insert the new entry:

309 UNICODE 5.2 see A6.7 *

</quote>

Not accepted

After FDAM3 was written, but before PDAM4 was created, the Unicode Consortium decided to align Unicode 5.1 with at least Amendment 3 and Amendment 4. This means that the FDAM3 text is no longer valid and indeed the text in PDAM4 reflects the current status. When FDAM3 is revised, it will be updated to remove all references to a Unicode 5.1 collection. The editor regrets the confusion.

E.2 pages 4-5 : Page 1357, Annex A.6 Unicode Collections

<quote>

At the end of Annex A.6, add new clause A.6.6 as follows.

A.6.6 308 UNICODE 5.1

308 The fixed collection UNICODE 5.2 consists of a fixed collection. The collection list is arranged by planes as follows.

</quote>

This should be:

<quote>

At the end of Annex A.6, add new clause A.6.7 as follows.

A.6.7 309 UNICODE 5.2

309 The fixed collection UNICODE 5.2 consists of a fixed collection. The collection list is arranged by planes as follows.

</quote>

At bottom of page 5

<quote>

NOTE - The collection 309 UNICODE 5.1 can also be determined by using another fixed collection from A.6 and several ranges of code positions.

Plane 00-10

Collection number and name

308 UNICODE 5.0

</quote>

should be:

<quote>

NOTE - The collection 309 UNICODE 5.2 can also be determined by using another fixed collection from A.6 and several ranges of code positions.

Plane 00-10

Collection number and name

308 UNICODE 5.1

</quote>

Not accepted

See disposition of comment E.1

E.3 page 25: Table 236 - Row F0: Mahjong Tiles

1F02A MAHJONG TILE JOKER

We suggest adding the Chinese name of this tile as an annotation:

1F02A MAHJONG TILE JOKER (baida)

Accepted

E.4 page 28-159: CJK Unified Ideographs Extension C

The codepoints for the following ranges of characters are printed with the last hexadecimal digit missing:

2AAAA..2AAAF printed as 2AAA (page 57)

2AABA..2AABF printed as 2AAB (page 57)

2AACA..2AACF printed as 2AAC (page 58)

2AADA..2AADF printed as 2AAD (page 58)

2AAEA..2AAEE printed as 2AAE (page 59)

2AAFA..2AAFD printed as 2AAF (page 59)

2ABAA..2ABAF printed as 2ABA (page 65)

2ABBA..2ABBF printed as 2ABB (page 65)

2ABCA..2ABCF printed as 2ABC (page 66)

2ABDA..2ABDF printed as 2ABD (page 66)

2ABEA..2ABEE printed as 2ABE (page 67)
2ABFA..2ABFD printed as 2ABF (page 67)
2ACAA..2ACAF printed as 2ACA (page 73)
2ACBA..2ACBF printed as 2ACB (page 73)
2ACCA..2ACCF printed as 2ACC (page 74)
2ACDA..2ACDF printed as 2ACD (page 74)
2ACEA..2ACEE printed as 2ACE (page 75)
2ACFA..2ACFD printed as 2ACF (page 75)
2ADAA..2ADAF printed as 2ADA (page 81)
2ADBA..2ADBF printed as 2ADB (page 81)
2ADCA..2ADCF printed as 2ADC (page 82)
2ADDA..2ADDF printed as 2ADD (page 82)
2ADEA..2ADEE printed as 2ADE (page 83)
2ADFA..2ADFD printed as 2ADF (page 83)
2AEAA..2AEAE printed as 2AEA (page 89)
2AEBA..2AEBE printed as 2AEB (page 89)
2AECA..2AECE printed as 2AEC (page 90)
2AEDA..2AEDE printed as 2AED (page 90)
2AEEA..2AEED printed as 2AEE (page 91)
2AFAA..2AFAD printed as 2AFA (page 97)
2AFBA..2AFBD printed as 2AFB (page 97)
2AFCA..2AFCD printed as 2AFC (page 98)
2AFDA..2AFDD printed as 2AFD (page 98)

Accepted

Similar comment from China and Japan

USA: Positive with comments:

Technical comments:

T.1 Character removal (Lanna)






The US is asking for the removal of the following characters:

1A65 LANNA VOWEL SIGN AM
1A66 LANNA VOWEL SIGN TALL AM





The rationale for their inclusion is provided in WG2 N3121:

The presence of [LANNA VOWEL SIGN] AM (and [LANNA VOWEL SIGN] TALL AM) follows the Thai convention of ensuring that a final consonant is not stored before the vowel it follows. This is the only situation in which it could occur and so [LANNA VOWEL SIGN] AM is encoded to alleviate the problem.



It is again clarified in WG2 N3207:

The written representation of /am/ involves two visual components:  VOWEL SIGN AA (or  VOWEL SIGN TALL AA) and  MAI KANG, which, if /am/ were not used, would be stored in that order (since final consonants are always stored after their vowels). In the case of /am/ the MAI KANG is often rendered as part of the preceding cluster to VOWEL SIGN AA. In order to ensure grapheme cluster integrity (see UAX#29 section 3) the unitary characters  and  for /am/ are proposed, following Thai practice. Note that /am/ is the only situation in which this occurs. The use of a sequence for AM would break the opportunity for a cluster boundary before AA. The characters may (if the UTC thinks it wise) be given compatibility decompositions to AA + MAI KANG and TALL AA + MAI KANG respectively. (In Thai, the decomposition for U+0E33 SARA AM is to 0E4D NIKHAHIT + U+0E32 SARA AA; this seems to be opposite, but Thai encodes in visual order so since the models are different this is not really relevant.)

The AM characters are an example of how sometimes more than one solution can be proposed for an encoding problem. It could be argued that these are “duplicate” characters, though the compatibility decomposition mitigates against that. One of the chief problems is that Northern Thai treats AM similarly to Thai AM; it places the MAI KANG glyph to the left of the -AA vowel (whether over the previous cluster or between the clusters):

, . In Khün and Lue, the MAI KANG render the MAI KANG over the -AA vowel: , . Without an encoded AM, it would be likely that Northern Thai users would confuse AA + MAI KANG and MAI KANG + AA, even though the latter is logically incorrect for the underlying phonemes. This is not a problem for Khün and Lue, which treat it as a vowel + final, but Northern Thai users think of it as equivalent to Thai AM.

Potentially, MAI KANG and AA may also occur with MAI KANG properly preceding AA, in different syllables.

For example /kam.wa:/ might be written  = KAL + MAI KANG + TONE-1 + SAKOT + WA + TALL AA while /kwa:m/ would be written  = KAL + SAKOT + WA + TONE-1 + TALL AM.

The explicitly-encoded AM gets around the problems of the re-ordering and ligation that would have to be solved if there were no AM, and would add a complexity that is not present in any of the surrounding scripts that contribute to the encoding milieu [sic] of the intended user community.

However all that long and detailed explanation does not remove the fact that these two characters are in fact equivalent to sequences of characters which are also proposed for encoding in the same document.

<U+1A63, U+1A76> for LANNA VOWEL SIGN AM, and
<U+1A64, U+1A76> for LANNA VOWEL SIGN TALL AM.

Proposing compatibility decomposition makes them even less useful as they will be filtered out by all processes using normalization form KC. It also makes them unsuitable for identifiers where the alternate sequences would be the only allowed representation.

In all cases, duplicate encoding should not happen in new proposals.

Propose acceptance

Similar request from UK (comment T1.1.3)

T.2 Addition of 2 Lanna characters

The US is also supporting the addition of the following Lanna characters as proposed by document WG2 N3207:

1A29 LANNA LETTER KHUN HIGH CHA
1AAD LANNA SIGN CAANG

Propose acceptance

Similar request from UK (comment T1.1.3)

T.3 Name and glyph changes for the new Cyrillic Extended-A

The US is in favor of the glyph and name changes as proposed in WG2 N3194 for the characters in the range U+2DE0..U+2DF5 (code position as originally presented in document SC2 N3914).

Propose acceptance

Similar request included in comment T2 from Ireland

T.4 Addition of 7 CJK Unified Ideographs

The US is also supporting the addition of 7 CJK Unified ideographs as proposed by document L2/07-67 (WG2 TBD) in positions U+9FBC through U+9FC2. At its last meeting, the IRG did not object to the fast-tracking of those characters, nor to their inclusion in Amendment 4. However, the IRG asked that those seven characters not be interleaved in Extension C, hence the proposed code points

These ideographs are present in the K-JIS and Sha-ken character collections. The K-JIS collection is developed by 共同通信社 and 配信先新聞社 for writing newspaper articles in Japan. The Sha-ken collection is part of a proprietary typesetting system widely used in Japan. These characters are also present in the Adobe-Japan1 collection, which is the basis for many desktop fonts, and at the time of this proposal are the only characters of that collection not present in Unicode / ISO/IEC 10646.

Glyph	Source collection	USource	Adobe-Japan1 CID	Proposed code point
壻	K-JIS #4431	UTC00836	15431	U+9FBC
燠	K-JIS #2191	UTC00835	15429	U+9FBD
燠	K-JIS #5304	UTC00837	15434	U+9FBE
萑	Sha-ken Index 7666	UTC00838	20068	U+9FBF
薺	Sha-ken Index 7614	UTC00839	20069	U+9FC0
誼	Sha-ken Index 7163	UTC00840	20070	U+9FC1
鶉	Sha-ken Index 7907	UTC00841	20071	U+9FC2

Propose acceptance

Table 1 CJK Unified Ideograph Extension C, consolidated table of issues

Code point	MB source	Sources	Current glyph	New glyph	Comment
2A719	Japan	V04-4059	破乙		? IRGN898 shows 破乙 for V04 source
2A71B	Japan	V04-405C	漱		? IRGN898 shows 漱 for V04 source
2A77D	China, Japan	V04-414D	倒酒	倒酒	Glyph change
2A825	China, Japan	V04-4272	去廣	去廣	Glyph change
2A8AB	China	TE-4633	聽	聽	Glyph change
2A949	Japan	V04-4536	廓	廓	Glyph change
2A988	UK	TC-553A	嫖		Remove, unification with 2177B 嫖
2AA11	Japan	V04-4652	嬪	嬪	Glyph change
2AA84	Japan	V04-473E	嗒	嗒	Glyph change
2AB23	China, Japan	V04-4839	慍	慍	Glyph change
2ABB1	China, Japan, UK	V04-4946	扞	掣	Glyph change (major)
2ABF6	China, Japan	V04-497B	散酒	散酒	Glyph change
2AC09	China	V04-4A25	縷	縷	Glyph change
2ACF3	China	JK-65923	啄	啄	Glyph change
2AE25	Japan	V04-4C78	濺		? IRGN898 shows 濺 for V04 source
2AEBE	Japan	V04-4D58	疇	疇	Glyph change
2AEC3	Japan	G_ZJW00576 V04-513D	釜	釜	Glyph change for V column, note that V04 uses 缶 instead of 父 as radical
2AEFF	UK	G_HC100898	獺		Remove, already encoded in 2F927, could add this source reference to it

2AF72	China	V04-4E31	底	底	Glyph change
2AF7C	Japan	TC-4D39 V04-4E34	吟 吟	吟	Glyph change for V column
2B000	China, Japan	V04-4F4B	砵	砵	Glyph change
2B028	China, Japan	V04-4F58	礪	礪	Glyph change
2B04B	China	TD-4223	禰	禰	Glyph change
2B09B	Japan	V04-5024	蔭	蔭	Glyph change
2B0CE	China, Japan, UK	V04-5035	筭	筭	Glyph change
2B151	Japan	G_HC501162 UTC00022	絶 珍		Sources seem correct, should these 2 sources been unified?
2B16C	Japan	G_ZFY00964 V04-513E	鋳 鋳	鋳	Glyph change for V column
2B187	China, Japan	V04-5142	甌 甌	甌	Glyph change
2B23C	Japan	TC-5F61 V04-5272	荅 荅	荅	Glyph change for V column
2B3BF	China, Japan	V04-554F	躡	躡	Glyph change
2B40E	Japan	V04-5478	躡	躡	Glyph change
2B642	Japan	V04-574C	馥	馥	Glyph change
2B644	Japan	V04-574E	馥	馥	Glyph change
2B70A	China	JK-66061	鵠	鵠	Glyph change

---end---